FINAL Functional Design Specifications Document

California Minimum Essential Datasets (MEDS)

10/23/2009

Michael Baker

Baker

Contents

Introduction

Purpose

This document defines the MEDS Functional Design Specifications based on the MEDS Functional Requirements (draft dated on September 22nd, 2009). It details the high level designs to meet the functional requirements and business rules. The approved functional design specifications shall form the basis for detailed Software Requirement Specifications (SRS) for proposed MEDS Data Map Services.

This document should not be confused with a detailed system SRS. Since this project is proceeding on an accelerated schedule, system design has to start while the Functional Requirements Document is still in the review process. This Functional Design Specification Document serves the purpose of informing the OCIO on how the functional requirements are intended to be met at a high level, so that the review process for the next phase, the SRS development, can be much smoother.

The stakeholders shall communicate any concerns and clarifications to the design team regarding the general design and technology direction as soon as possible.

Scope

This document is intended to provide an overview of methodologies and technologies to be used to meet the functional requirements captured in the MEDS Functional Requirements Document.

The technical overview defines the overall business process and should enable business level users to understand the system in a broader prospective.

The use cases mentioned below capture the specific functional and business rules applied on each independent business process activity and should enable the system users to validate the functionalities supported implicitly and explicitly by the system. The use cases also identify the end users of the system.

Abbreviations/Definitions

Term	Definition
Actor	An actor in the Unified Modeling Language (UML) "specifies a role played by a user or any other system that interacts with the subject.
AJAX	Shorthand for asynchronous JavaScript and XML), is a group of interrelated web development techniques used on the client-side to create interactive web applications or rich Internet applications.
API	Application Programming Interface
Cloud Computing	A paradigm of computing in which dynamically scalable and often virtualized resources are provided as a service over the Internet
Crowd Sourcing	A neologism for the act of taking tasks traditionally performed by an employee or contractor, and outsourcing it to an undefined, generally large group of people or community in the form of an open call
CSW	Web Catalog Service: Access to catalog information

DHS	Department of Homeland Security
EOC	Emergency Operations Centers
EROS Center	USGS Earth Resources Observation and Science Center
ETL	Extract, Transform, and Load
FME	A GIS ETL tool by SAFE Software
FTP	File Transfer Protocol
GEORSS	An emerging standard for encoding location as part of a Web feed.
GeoXACML	Geospatial eXtensible Access Control Markup Language (currently being standardized)
GML	Geography Markup Language: XML format for geographical information.
GUI	Graphical User Interface
HDDS	USGS Hazards Data Distribution System
HTTP	Hyper Text Transfer Protocol
KML	Keyhole Markup Language
MEDS	Minimum Essential Data Set
OGC	Open Geospatial Consortium
RSS	Really Simple Syndication
SDSC	San Diego Super Computer Center
SFS	Simple Features – SQL
SLA	Service Level Agreement
SRS	Software Requirement Specification
SMS	Short Message Service
UML	Unified Modeling Language
WCS	Web Coverage Service: Provide coverage objects from a specified region.
Web 2.0	Commonly associated with web development and web design that facilitates interactive information sharing, interoperability, user-centered design and collaboration on the World Wide Web.
WFS	Web Feature Services
WFT	Web Feature Transactions
WMS	Web Mapping Services
WPS	Web Processing Service: Remote processing service

Inputs

MEDS Final Functional Requirements (dated October 2nd, 2009).

MEDS stakeholder meeting notes.

MEDS 09/14/09 Steering Committee meeting notes.

Technical Overview

The primary objectives of the proposed MEDS solutions is to provide a scalable, high performance and maintainable platform for hosting and distributing MEDS data, and also to establish a framework for data providers to contribute to future data improvement and updates. Specifically:

1. Provide secure and reliable hosting of MEDS data.
2. Provide reliable and speedy access to the MEDS data via Web Services, file download, and sneakernet.
3. Provide a technical framework based on open standards that can accommodate additional web service development on top of MEDS data in the future.
4. Provide a method for stakeholders to send data updates to MEDS.

User Characteristics

Users of MEDS data and Map Services are government GIS professionals or technical staff who should have a working knowledge of GIS and spatial databases. Initially MEDS users are emergency response personnel (Emergency Operations Centers (EOC)).

Actor	Description
Metropolitan EOCs (User)	It is anticipated that urban, well-funded EOCs with adequate internal IT infrastructure will prefer to rely on data stored within their own networks in order to minimize disruption in case of external connectivity failure. Primary benefits of MEDS for this type of client will be as a backup repository for their own local data to be used for disaster recovery and as an easily accessed source of data for adjacent jurisdictions. Normal Use Daily use would consist of checking MEDS for updates and downloading updated datasets for adjacent jurisdictions. Uploading new data from local sources to MEDS would be considered routine maintenance. Normal use might also include running HAZUS and other models, developing evacuation plans and other pre-planning activities. Event Use Assuming that the EOC has up-to-date data when an event occurs, their event-driven usage of MEDS may emphasize access to post-event data. Download of or service-based access to post-event imagery could be a critical use. Imagery could be loaded once and used by many agencies. Pushing high-priority event-based local data up to MEDS in order to share the common operating picture (COP) with other local agencies, adjacent jurisdictions and state agencies could be beneficial.

Rural EOCs (User)	EOCs that lack internal capacity for data storage or processing may see benefit in using MEDS services more directly for data storage, access and applications. Normal Use Daily use would consist of accessing the MEDS data sets for disaster planning activities. Uploading new data from local sources to MEDS would be considered routine maintenance. Event Use Adequate bandwidth is a concern for some rural jurisdictions. A MEDS "data pack" consisting of hard drives with relevant MEDS data that can be rapidly deployed to responding agencies could help mitigate this concern. All EOCs will have a need to create maps for distribution to field crews, public officials, and the press. MEDS data layers can serve as a common, easily accessed backdrop.
CalEMA State Offices (User)	One primary benefit of MEDS to state-level agencies will be the pervasive availability of easily accessed data sets. This should make the construction of common operating pictures significantly easier across organizational boundaries. Normal Use MEDS data will be used for disaster planning activities. Event Use During and immediately after an event, CalEMA staff can use MEDS data in combination with other datasets to assist with disaster assessment and recovery. It is anticipated that CalEMA staff will use the web-based data services to access MEDS, rather than needing a local copy. CalEMA staff may play a role in pushing post-event data (event-related landmarks, for instance) into MEDS.
State/Federal Agencies (User)	Normal Use Routine pushes from data stewards (e.g., Caltrans, USGS) will be required to keep some data sets updated. Development of relevant functionality, such as buffer analysis of landmarks (e.g., query all schools within 1 mile of an event location) could lead to routine usage of MEDS data. This would be true for county and regional level normal use as well. Event Use Post-event data pushes will be required, especially for imagery. State, regional, and federal agencies may be able to use MEDS landmarks to enhance their COPs.

General Constraints and Risks

1. MEDS Web Services require constant and reliable internet connectivity and some data services demand broad bandwidth. Some rural areas do not have reliable internet connectivity and bandwidth is limited. Any system design shall consider this constraint and provide alternatives.
2. The majority of data stakeholders and professional GIS users of the targeted government agencies use ESRI technologies. Tremendous investment has been made in the ESRI technologies in creating and maintaining spatial data, providing data analysis and visualization functions. Most staff's GIS training has also been ESRI-centric. It is important to consider this constraint factor in architectural and application designs.
3. MEDS map services have a high level of up-time requirement. They are required to be 24x7, even though no specific SLA (Service Level Agreement) up-time is defined. Cost is an obvious constraint factor. Any data storage and web service design needs to consider the high SLA requirement.

Assumptions and Dependencies

1. It is assumed that two core data centers will be available for data storage and web servicing hosting: NASA Ames Research Center and San Diego Supercomputer Center.
2. It is assumed that a number of mirrored sites would be available for MEDS data downloads: Cal-Atlas may be an appropriate mirror site for downloads. Other potential mirror sites might include the Hazard Data Distribution System (HDDS), NIFC, Geomac, and others.
3. It is assumed that there is a steward for each of the MEDS dataset. Per the Steering Committee Meeting Notes, CalEMA will be the Landmarks steward and Caltrans will be the Transportation steward.
4. While the data stewards might also serve as hosts for their MEDS data, this is not a required role. A more robust solution may be for the stewards to simply maintain data and to regularly push their data to the core data centers for distribution. Benefits of this approach include taking advantage of the up-time, performance and bandwidth of the core data centers, reducing load on the stewards' infrastructure, and simplifying security and administration of the MEDS.

Data Service Applications

1. Data Synchronization and Replication

There shall be a well defined data schema between the MEDS and data providers to minimize efforts required for data synchronization and replications. Ideally these processes shall not require extensive human interventions.

Since most data providers use ArcSDE to store their spatial data, it is reasonable to use ESRI synchronization and replication functions. For additional data ETL (Extract, Transform, Load) needs, off-the-shelf tools such as FME can be utilized.

When the data is synched up to MEDS, there shall be a data instance specific for QA/QC purposes. Data synchronization shall not occur directly between the production MEDS data instance and the data provider instance.

2. Data Change Notification

There shall be a push notification system to notify users of data edits and changes. This shall be a push process and technical support staff shall decide what and when to notify the users. The methods of the push shall include email notification and RSS feeds.

There shall be an automated process to record the MEDS data changes and this change log shall form the foundation for push notifications.

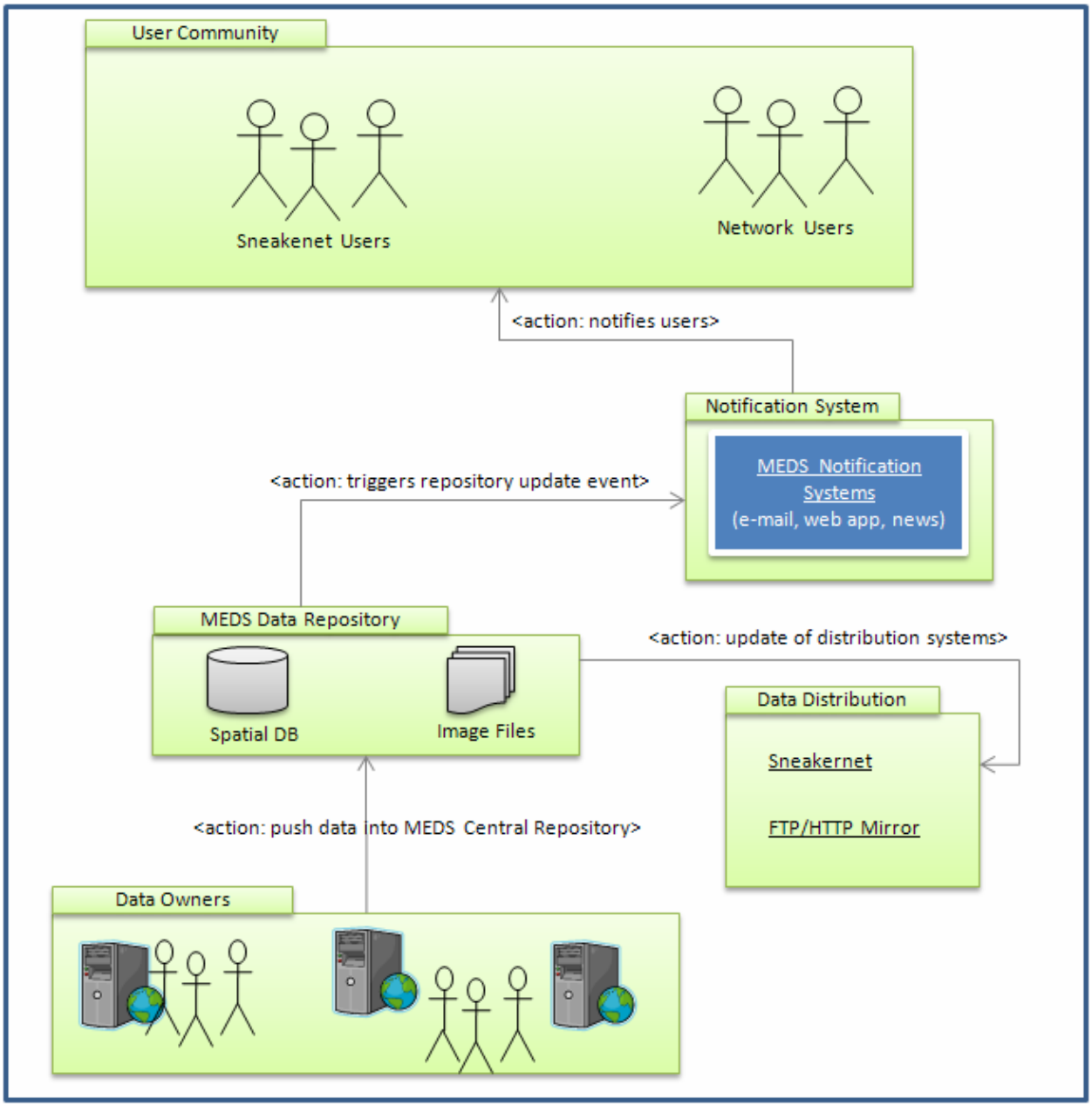

Figure 1. Data Synchronization, Replication Change Notification

3. Web Mapping Services

Five types of web mapping services shall be created to serve the MEDS data and also promote the presence of the MEDS. All mapping services shall be OGC compliant. Figure 2 illustrates a typical architectural framework of data storage, distribution, and services based on OGC standards (http://en.wikipedia.org/wiki/File:Geoservices_server_with_apps.png). The five web services include: WMS, WFS, GeoRSS, KML, and WCS.

WFS can slow down the overall server performance if not implemented properly, for example, by allowing viewing or editing of very large datasets. MEDS WFS shall be provided only at a smaller geographic scale while WMS provides visualization at all geographic scales.

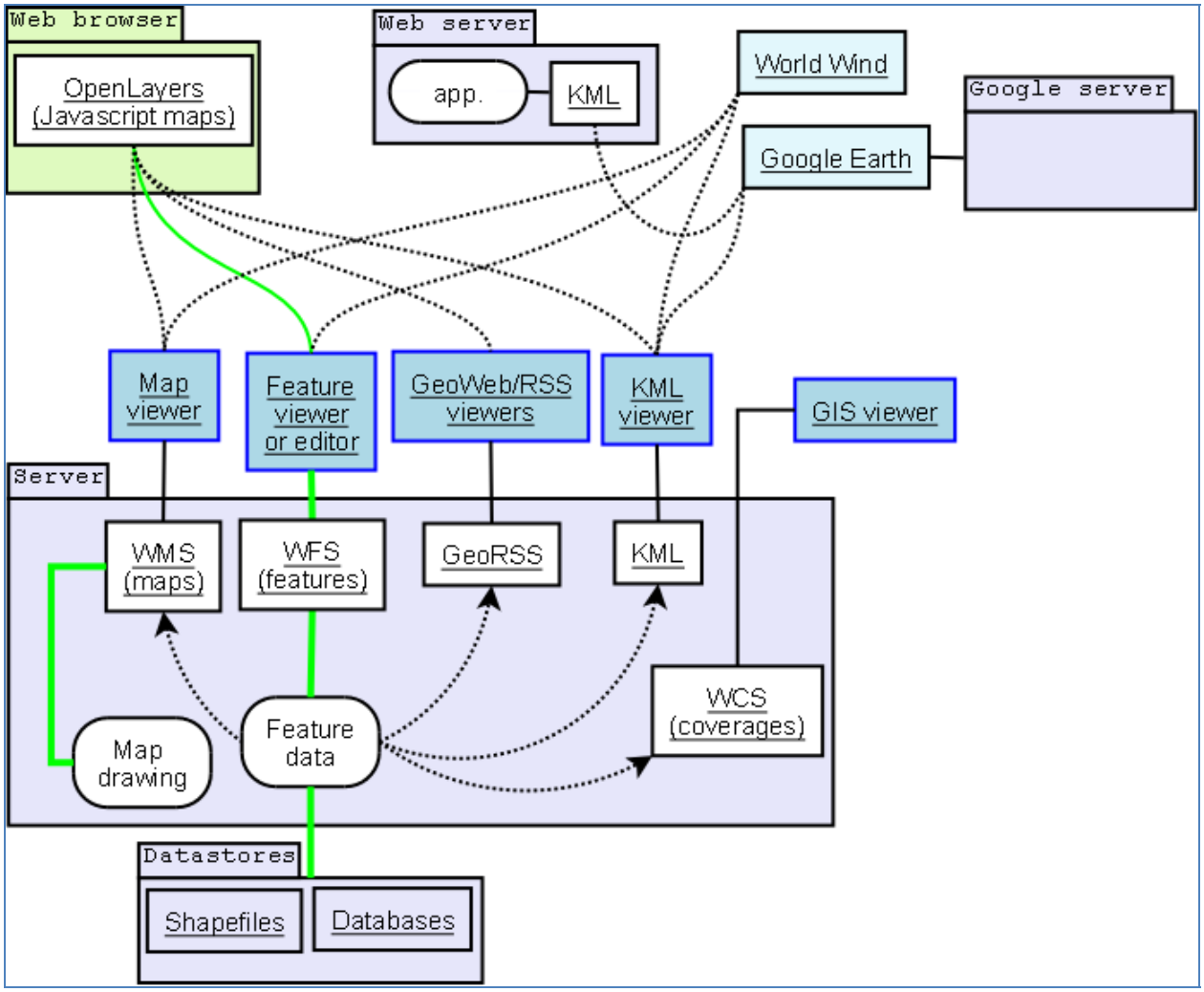

Figure 2. OGC Geospatial Web Service Framework

Data load balancing for web services shall be a requirement for system design. There shall be at least two data centers hosting both data and applications for fail-safe purposes and for added scalability and performance.

There shall be a process to notify the consumers of the web services in case of catastrophic failure of either or both data centers. A monitoring application shall be developed and hosted outside these two data centers to check on the system down-time and send emails, RSS feeds and SMSs to both technical support staff and web service users.

There shall be detailed web documentation on how to use the web services. Sample projects should be provided. MEDS shall also provide an API Playground, modeled after Google API and Microsoft API playgrounds. Users shall be able to edit sample code interfacing with the APIs and see how they work. Users shall also be able to save and copy/paste the source code.

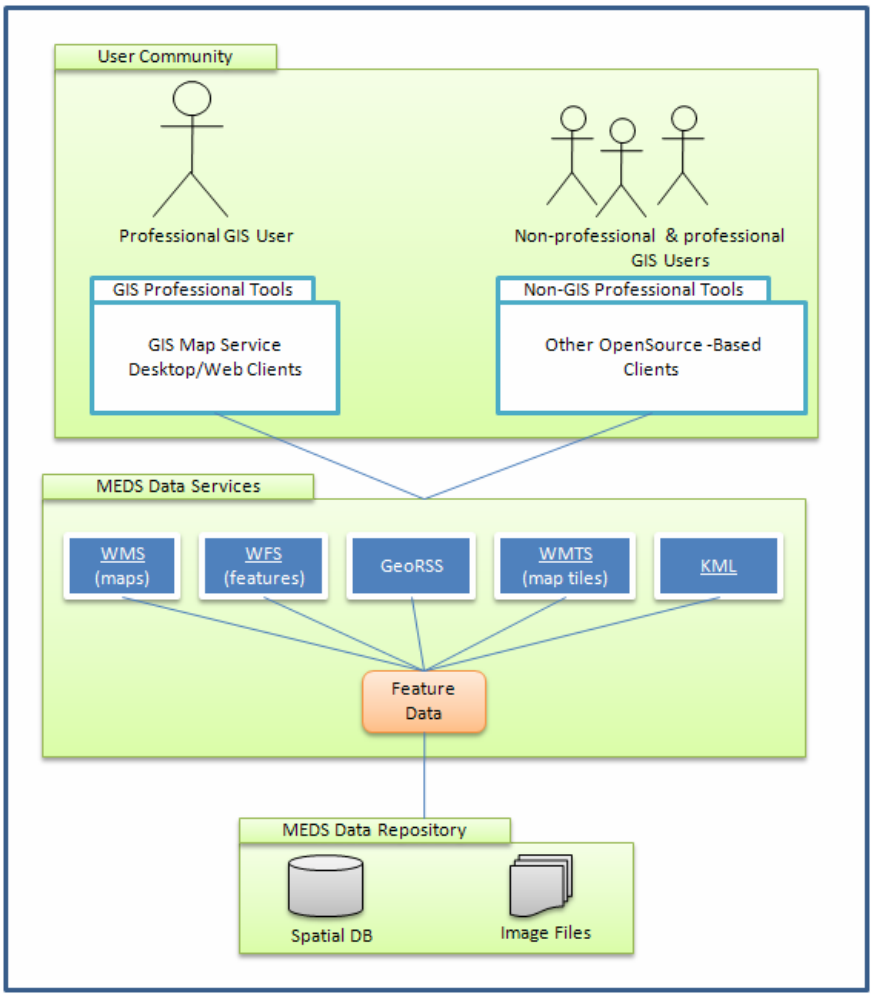

Figure 3. MEDS Web Service Framework

4. Web 2.0 APIs to Serve MEDS Imagery in Cached Tiles

Even though early MEDS users will be GIS professionals at government agencies, who probably would use WMS for imagery in their professional GIS client tools, the need for non-professional users to visualize MEDS images on popular mapping engines such as Google Maps and Bing Maps are on the rise and they may demand MEDS to provide such service. Professional users may also have the need to use the images in this way for non-mission critical uses and faster loading time.

A set of tiles of the MEDS images could be created in Geographic Coordinate System, hosted on the data centers, and made available via JavaScript. Users do not need professional client tools to overlay those cached tiles on Google Maps or Bing Maps.

Those cached tiles shall also be included in the Sneakernet distribution. Optionally this set of tiles can also be offered to commercial hosting services such as Google and Microsoft.

The sample code for using the cached tiles shall also be made available in the Code Playground previously discussed. The sample code shall demonstrate how MEDS tiles can be used in all major online mapping engines such as Google Maps, Bing Maps, and Yahoo Maps. Special efforts shall be made to demonstrate the uses of MEDS tiles on Open Source engines such as OpenLayers.

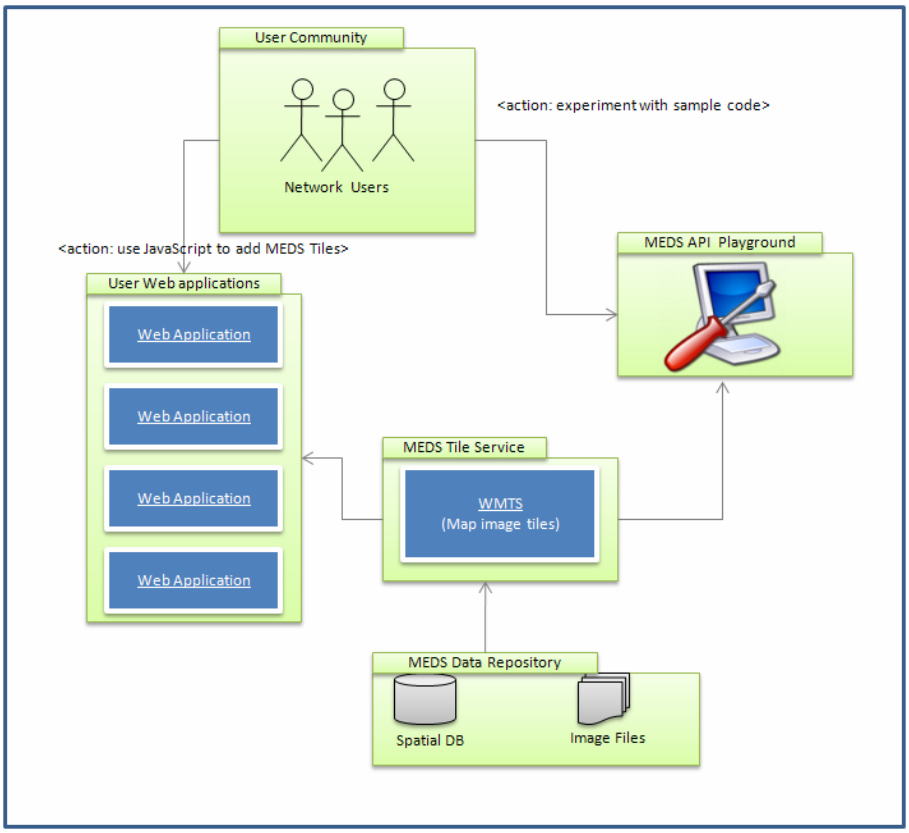

Figure 4. MEDS Web 2.0 Image Service

Data Catalog and Distributions

MEDS shall have reliable data storage, fast distribution, and be easy to discover.

1. Data and File Storage

MEDS includes primarily spatial datasets, both vector and raster. All MEDS vector data are provided as ESRI formats or one of the formats that can be translated into an ESRI format by the data owners. While most of images (aerial and/or satellite photos) initially are provided as files, pyramided files or tiles can be created to enable faster loading in most mainstream GIS and mapping tools.

Vector data storage shall be in an enterprise spatial database with versioning capability regardless of how the source data is provided. This spatial database shall serve as the official MEDS database for the vector data. A set of GIS files such as shape file, Personal GeoDatabases, or KML files shall be created and zipped by county level from this spatial database.

Images shall be stored at the highest resolution and lowest possible compression as geo-referenced image files. There shall be a mechanism to record the metadata of each image. A set of pyramided image files can be created for a smaller geography, and a set of lower resolution of images can be created for quick file loading.

A set of pre-defined database and file packages shall be developed to reduce both distribution confusion and distribution size. For the vector data, files shall be created at the county level and can be aggregated to create regional datasets. For the image data, files shall also be organized at the county level

The official storage of data and files shall be on at least two data centers for fail-safe purpose.

2. Data and File Distribution
    a. FTP/HTTP Distribution

        FTP is the preferred method for large dataset and file download due to its better transfer reliability over HTTP download method. For smaller dataset and files HTTP shall also be supported.

        In addition to a minimum of two core data centers, there shall be multiple mirrored FTP and HTTP download sites in California to increase the download reliability and speed. There shall be at least three additional FTP and HTTP mirrored sites in the state on reputable data centers such as those operated by large government agencies and universities. The OCIO shall encourage additional mirrored service while maintaining a list of endorsed sites.

        The mirrored sites shall also consider the download needs from Central and Eastern US (for Federal access). MEDS shall explorer the possibility of leveraging existing governmental data hosting infrastructure such as DHS and Eros Data Center via HDDS.

    b. Sneakernet Distribution

High speed external storage devices (USB 2.0 and above) shall be the preferred sneakernet distribution method. A set of such devices shall be made available for each pre-defined data packages. Three data distribution centers shall be created in Northern, Central, and Southern California to reduce emergency situation delivery time. A complete set of data shall also be available in the Central and Eastern US.

3. Data Discovery

MEDS data discovery shall be made easy via proper channels such as Web page presence, RSS and GEORSS feeds, and email list. Most popular geospatial data aggregators such as Geospatial OneStop and HDDS can consume and publish standard RSS and GEORSS feeds.

Some popular GIS client software tools have built-in data discovery functions, and MEDS shall make it easy to be found via those functions.

4. Use Cases

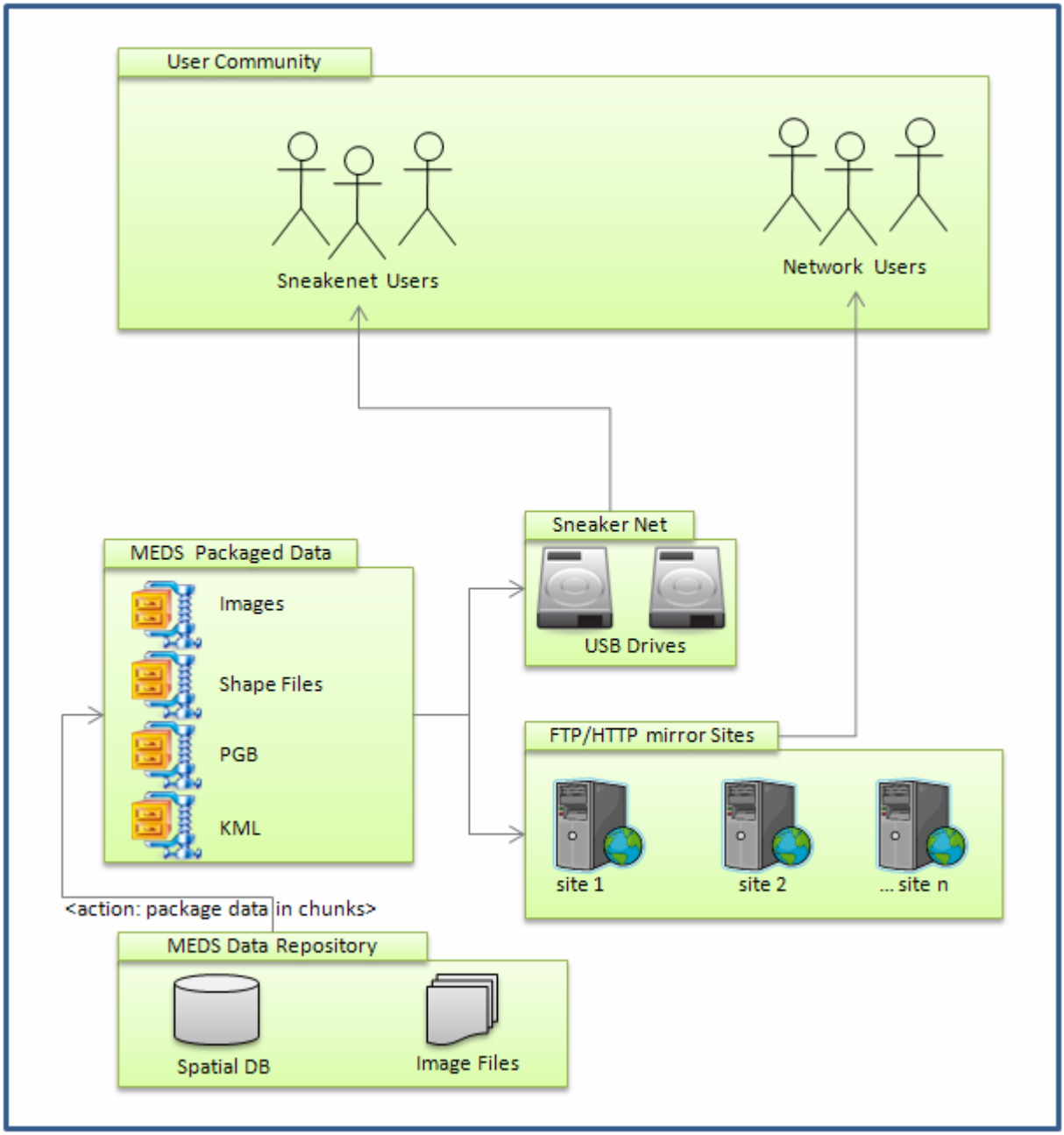

Figure 5. Data Catalog and Distribution

Security

Initial users of MEDS are government agencies, and both data and mapping services shall be provided in a secured way. All MEDS database and application activities shall be in compliance with hosting service security requirements.

From a data and mapping services perspective, MEDS shall not require an additional security layer and can exist in the public web server layer and shall be read-only.

From data and API access perspective, a simple security measure shall be implemented that checks user name, password, and domain name/IP address. No additional security shall be needed since all data in MEDS can be obtained publically. The security measure is provided with the objective of restricting access in order to provide better performance for the selected group.

Implied Requirements

1. User friendly GUI for any end –user applications.
2. OGC compliance in Mapping Services.
3. Metadata available for all MEDS datasets and in compliance with FGDC.
4. All users shall have FTP software when and if they access to the MEDS data via FTP software.

Requirement Validation

This functional design specification document is based on the Final Functional Requirement Document dated October 2nd, 2009. Changes in the requirements would have impacts on this document. Both documents shall be updated and stay in sync.

Traceability

Version	Description	Changed By	Date
1.5	DRAFT as delivered to OCIO	Baker	9/29/2009
1.5_clo_comments	OCIO comments on DRAFT	OCIO	10/19/2009
2.1	FINAL document version, incorporating OCIO comments	Jay, Bud	10/23/2009
2.2	FINAL document version, incorporating OCIO comments. All comments and edits incorporated/accepted.	Jay, Bud	10/23/2009